VU Research Portal

Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic Escherichia coli

Otto, B.R.; Sijbrandi, R.; Luirink, S.; Oudega, B.; Heddle, J.G.; Mizutani, K.; Park, S.Y.; Tame, J.R.H.

published in
Journal of Biological Chemistry
2005

DOI (link to publisher)
10.1074/jbc.M412885200

document version
Publisher's PDF, also known as Version of record

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 17, Issue of April 29, pp. 17339–17345, 2005
Printed in U.S.A.

Crystal Structure of Hemoglobin Protease, a Heme Binding Autotransporter Protein from Pathogenic *Escherichia coli**

Received for publication, November 15, 2004, and in revised form, February 17, 2005
Published, JBC Papers in Press, February 22, 2005, DOI 10.1074/jbc.M412885200

Ben R. Otto‡, Robert Sijbrandi‡, Joen Luirink‡, Bauke Oudega‡, Jonathan G. Heddle§, Kenji Mizutani§, Sam-Yong Park§, and Jeremy R. H. Tame§¶

From the ‡Department of Molecular Microbiology, Faculty of Earth and Life Sciences, Vrije Universiteit de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands and the §Protein Design Laboratory, Yokohama City University, Suehiro-cho 1-7-29, Tsurumi, Yokohama 230-0045, Japan

The acquisition of iron is essential for the survival of pathogenic bacteria, which have consequently evolved a wide variety of uptake systems to extract iron and heme from host proteins such as hemoglobin. Hemoglobin protease (Hbp) was discovered as a factor involved in the symbiosis of pathogenic *Escherichia coli* and *Bacteroides fragilis*, which cause intra-abdominal abscesses. Released from *E. coli*, this serine protease autotransporter degrades hemoglobin and delivers heme to both bacterial species. The crystal structure of the complete passenger domain of Hbp (110 kDa) is presented, which is the first structure from this class of serine proteases and the largest parallel β-helical structure yet solved.

Autotransporters (ATs)[1] are virulence proteins produced by a variety of pathogenic bacteria, including several types of *Escherichia coli* associated with severe, even fatal, diarrhea and urinary tract infections. An increasing number of these proteins have been discovered over the last decade, and a large body of research is now devoted to these proteins and their different roles in various diseases. ATs have evolved a unique export mechanism, the type V pathway (1, 2). Initially the protein is synthesized with an N-terminal leader peptide directing the protein to the periplasm via the signal recognition particle (SRP) (3). Once across the inner membrane, the C-terminal domain forms a β-barrel structure within the outer membrane, allowing the passenger domain to escape to the external medium, where it is released proteolytically from the cell (4–6). Recently the crystal structure of a translocator (membrane) domain of an AT has been solved (7). AT passenger domains are generally more than 100 kDa in size and vary widely in sequence and function but include a group of closely related serine proteases, the SPATE family (serine protease autotransporters of the Enterobacteriaceae). These proteins are found only in pathogenic bacteria, but their precise roles remain unknown (8, 9). The SPATE family includes EspC from enteropathogenic *E. coli* (10), EspP from enterohemorrhagic *E. coli* (11), plasmid-encoded toxin (Pet) from enteroaggregative *E. coli* (12), Sat from uropathogenic *E. coli* (13), Tsh and Vat from avian pathogenic *E. coli* (14, 15), and Pic and SepA from *Shigella flexneri* (16, 17). These AT proteins are associated with a number of serious, possibly life-threatening diseases affecting millions world-wide. Tsh and Vat are essential virulence factors in avian diseases that cost the poultry industry millions of dollars annually. SPATE proteins do not appear to have conserved or even related functions (8, 9). Vat has vacuolating activity, and Pet enters eukaryotic cells to disturb the cytoskeletal system. As essential proteins in pathogenesis, SPATE proteins are key drug targets and potential routes to vaccines. Knowledge of the molecular structure is a first step toward understanding their mechanisms and hopefully how they may be inactivated. We have solved the crystal structure of the passenger domain of heme binding hemoglobin protease (Hbp), an AT involved in bacterial synergy in the development of peritonitis.

Peritonitis is an inflammation of the peritoneal cavity and may develop into a life-threatening condition through the formation of persistent abscesses, which leak pathogenic bacteria into the general circulation (18, 19). This common complication of invasive surgery then results in septic shock and multiorgan failure with mortality rates well in excess of 50%. These abscesses are difficult to treat with antimicrobial therapy, and usually require surgical intervention. It is frequently found that pathogenic *E. coli* and the strictly anaerobic *Bacteroides fragilis* occur together in intra-abdominal abscesses, the combination surviving better than either species alone (20, 21). *B. fragilis* is a common member of the normal human gut flora, but it is also a significant opportunistic pathogen (18, 19). It promotes fibrin deposition by the host, which localizes the bacteria and prevents systemic spread, but also provides an environment where the bacteria can escape attack by the immune system. In return, *E. coli* helps to supply both species with heme scavenged from hemoglobin (Hb). Hbp was discovered as the protease secreted by *E. coli*, which degrades Hb and binds heme (21, 22). The *hbp* gene is located on a large episome, pColV-K30, which also encodes an aerobactin-dependent iron uptake system (23). Hbp is highly expressed by *E. coli* during peritonitis and plays a key role in bacterial synergy with *B. fragilis*. Adding Hbp and heme to cultures of *B. fragilis* was shown to stimulate growth, and mice immunized with Hbp showed no abscess formation when subsequently challenged with mixed *E. coli*/*B. fragilis* cultures. Hbp therefore appears to be a useful target in the treatment and prevention of peritonitis. We have determined the crystal structure of Hbp with the longer term view of producing inhibitors, vaccines, or other treatments for peritonitis and other diseases caused by SPATE proteins. From this first molecular model of an AT, we have

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1WXR) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ To whom correspondence should be addressed. Tel.: 81-(0)45-508-7228; Fax: 81-(0)45-508-7366; E-mail: jtame@tsurumi.yokohama-cu.ac.jp.

[1] The abbreviations used are: AT, autotransporter; Hb, hemoglobin; Hbp, hemoglobin protease.

TABLE I
Data collection and refinement statistics

Space group, unit cell dimensions (Å)	*P*6122, $a = b = 114.86$, $c = 437.05$		
Wavelength (Å)	Peak	Inflection	Remote
Resolution range (Å)	50.0–3.0	50.0–3.0	50.0–3.0
Reflections (measured/unique)	214,532/33,539	242,451/34,130	232,344/33,910
Completeness (overall/outer shell, %)[a]	95.4/88.5	98.8/92.6	96.0/87.7
R_{merge} (overall/outer shell, %)[a,b]	10.2/21.2	8.3/18.3	8.5/18.8
Redundancy (overall)	6.4	7.1	6.9
Mean $\langle I/\sigma(I)\rangle$ (overall)	7.8	8.9	8.3
Phasing (20.0–3.0 Å)			
Mean FOM[c] after SOLVE phasing	0.33		
Refinement statistics			
Refinement resolution (Å)	30.0–2.20		
Reflections (measured/unique)	525,426/81,279		
Completeness (overall/outer shell, %)[a]	92.3/70.8		
R_{merge} (overall/outer shell, %)[a,b]	5.8/31.8		
Redundancy (overall)	6.5		
Mean $\langle I/\sigma(I)\rangle$ (overall)	11.0		
σ cut-off /reflections used	0.0/81,279		
R-factor[d]/free *R*-factor (%)[b]	20.3/24.3		
Root mean square deviation bond lengths/bond angles (Å)	0.016/1.7		
Water molecules	399		
Average *B*-factor (protein/water, Å^2)	35/33		
Ramachandran plot			
Residues in most favorable regions (%)	86.1		
Residues in additionally allowed regions (%)	12.9		
Residues in generously allowed regions (%)	1.0		

[a] Completeness and R_{merge}, are given for overall data and for the highest resolution shell (overall/the highest resolution shell). The highest resolution shells for the MAD data and native data are 3.11–3.0 Å and 2.28–2.20 Å, respectively.
[b] $R_{merge} = \Sigma I_i - \langle I\rangle / \Sigma I_i$; where I_i is intensity of observation and $\langle I\rangle$ is the mean value for that reflection.
[c] Figure of merit (FOM) = F_{best}/F .
[d] *R* factor = $\Sigma \| F_o(h) - F_c(h) \| / \Sigma h F_o(h)$; where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. The free *R*-factor was calculated with 5% of the data excluded from the refinement.

constructed homology models of other ATs to observe both common and unique features of these proteins.

MATERIALS AND METHODS

Protein Purification and Crystallization—The cloning of Hbp has been described previously (22). *E. coli* strain DH5α harboring plasmid pACYCHbp was grown to an A_{600} of 0.6 in MA medium, and the supernatant was collected and concentrated. Following 1000-fold concentration by ultrafiltration, the protein was purified by gel filtration in 10 mM Tris-Cl, pH 8.0, 300 mM sodium chloride. Apo-Hbp was crystallized by the published method (24). 4 mg/ml protein was crystallized using the hanging drop method using 0.1 M sodium acetate, pH 4.6, 200 mM ammonium sulfate, 10% glycerol, and 15–20% polyethylene glycol 6000.

Data Collection and Processing—Multi-wavelength x-ray data were collected at beam line 45XU-PX, SPring8, Harima, Japan. Native data were collected at beam line BL5 of the Photon Factory, Tsukuba, to a resolution of 2.2 Å. All data were indexed and integrated using HKL2000 and scaled with the program SCALEPACK (25). The crystals are in space group *P*6122 with unit cell dimensions of $a = b = 114.86$ Å, $c = 437.05$ Å, with one molecule per asymmetric unit and a solvent content of 65%.

Structure Determination and Refinement—The structure of Hbp was solved by MAD (multiple-wavelength anomalous dispersion) using selenomethionine-containing protein. Phase determination and refinement were performed using the programs SOLVE and RESOLVE (26, 27). 12 Se sites were located out of 14 expected. The 3.0-Å resolution electron density map obtained from RESOLVE was of sufficient quality to permit unambiguous chain tracing without further refinement of the heavy atom sites. After model building using the programs O (28) and Turbo (29), initial refinement was carried out with XPLOR (30) and final refinement with REFMAC (31, 32) using the 2.2-Å resolution native data. Solvent molecules were placed at positions where spherical electron-density peaks were found above 1.5 σ in the $2F_o - F_c$ map and above 3.0 σ in the $F_o - F_c$ map and where stereochemically reasonable hydrogen bonds were allowed. The final model extends from the N terminus (residue 1 of the mature protein) to residue 1048, the C-terminal residue of the passenger domain. The Ramachandran plot has 86% of the residues in the most favorable regions and none in disallowed regions. A summary of the data collection and refinement statistics is given in Table I. The final model and structure factors have been deposited in the Protein Data Bank (PDB, 1WXR).

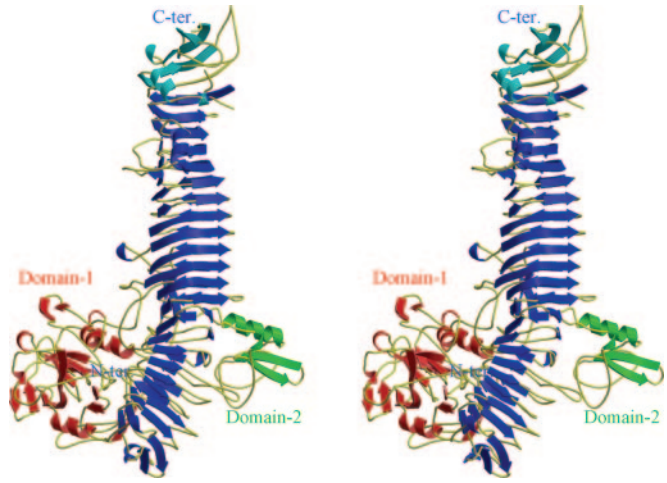

FIG. 1. **A stereo ribbon diagram showing the overall structure of the entire passenger region of Hbp.** Helices and strands are colored *red* for the protease domain (domain 1, residues 1–256) and *green* for the chitinase b-like domain (domain 2, residues 481–556). The β strands in the parallel β helix are colored *dark blue*. Residues 950–1048 (the C terminus of the mature protein) are colored *light blue*. This region is required for transport of the passenger domain through the membrane pore created by the C-terminal region of the complete protein.

RESULTS

The overall structure of the protein is shown in Fig. 1. The most striking feature is the right-handed β-helical C-terminal tail, which extends from residue 260 onwards, forming three parallel β sheets. This is highly similar to pertactin (PDB 1DAB) (33) and the N terminus of filamentous hemagglutinin (PDB 1RWR) (34), both adhesins from *Bordetella pertussis*. Pertactin is the only AT structure refined to date, one of only a dozen or so right-handed parallel β helix protein structures so far solved. The first was the pectate lyase PelC (PDB 1AIR),

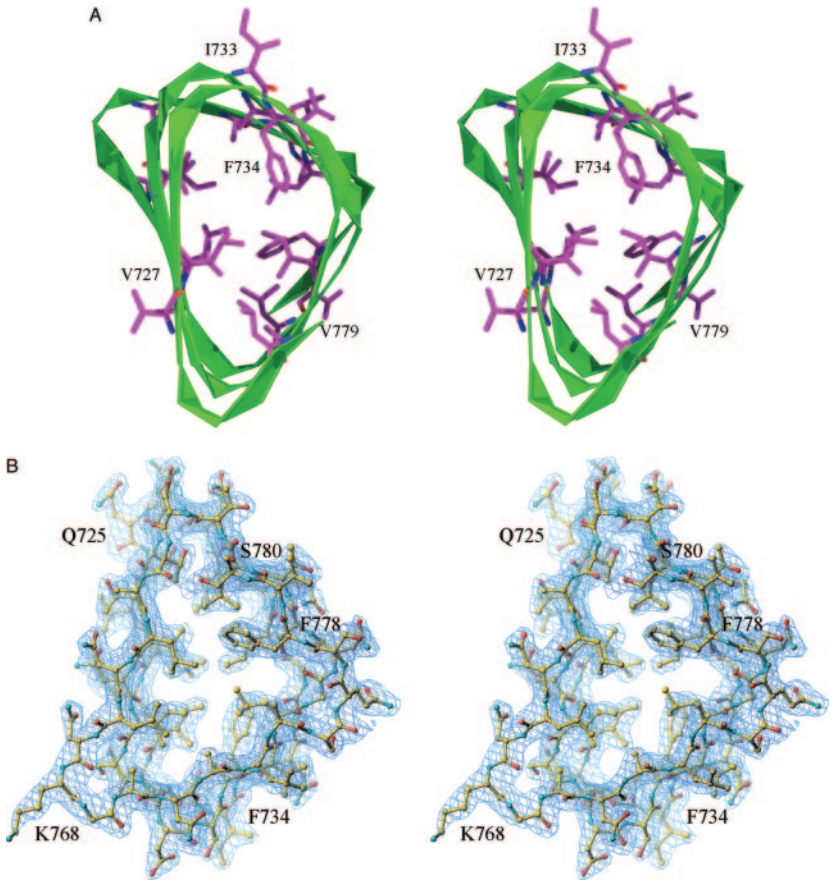

FIG. 2. *Top*, a slice through the β helix showing the hydrophobic core. The backbone is shown as a *green ribbon*, and hydrophobic side chains are shown in *pink*. *Bottom*, the $2mF_o - DF_c$ electron density map, shown in stereo and contoured at 1.3σ, showing a cross-section through the β helix.

and a number of different pectate lyase structures are now known (35). The pectate lyase superfamily members are a functionally eclectic mix, including P22 tailspike protein (PDB 1TYU), chondroitinase B (PDB 1DGB), and insect cysteine-rich antifreeze protein (PDB 1EZG). All share a similar pattern of repeats of about 25 amino acid residues, which form one turn of the right-handed helix. Each turn creates three β strands that form extended β sheets with neighboring turns. The core of this region is generally but not exclusively hydrophobic. The leucine-rich repeat family is another group of protein structures formed from repeated sequence motifs, but these include an α helix in each turn and are found to be highly curved into horseshoe shapes. The pectate lyase superfamily forms a much straighter superhelix. The shape of Hbp was previously determined to low resolution using a combination of small-angle x-ray scattering and analytical ultracentrifugation (36). The crystal structure shows a slightly straighter structure than the model from small-angle x-ray scattering, possibly due to greater flexibility in solution.

The Hbp model includes 24 turns of right-handed β helix, compared with 16 in pertactin, the longest parallel β-helix known to date. Hbp is highly reminiscent of pertactin, but less regular and slightly kinked in the center. It also has a large N-terminal globular domain and a smaller domain of 75 residues (481–556) about half-way along the β-helix, as well as decorative loops where the polypeptide chain departs from the helical stem. The extra surface features make the length of Hbp (1048 residues) almost twice that of pertactin (539 residues). The extended hydrophobic core is largely but not exclusively filled with aliphatic side chains (Fig. 2, *top* and *bottom*). The N-terminal region close to the serine protease domain has a higher proportion of buried aromatic residues. Very few polar residues point into the hydrophobic core, but Ser^{780} is one exception (Fig. 2, *bottom*). Other SPATE proteins generally

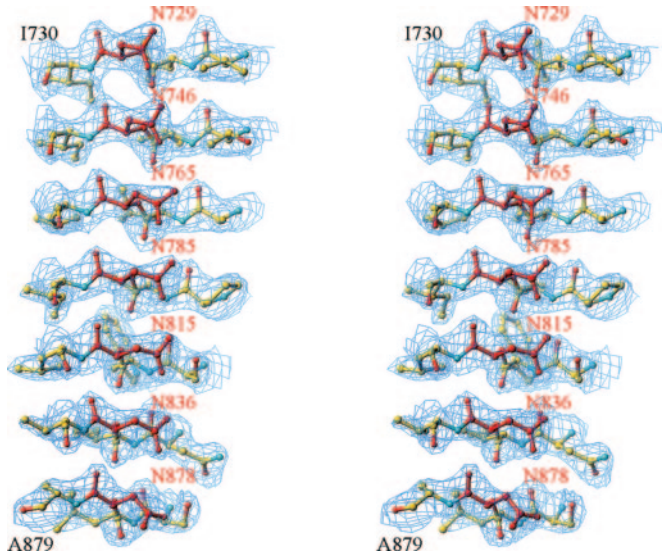

FIG. 3. **An orthogonal view to Fig. 2 (*bottom*) of the final $2mF_o - DF_c$ electron density map, at the same contour level, showing a stack of asparagine residues on the surface of the β helix (colored *red*), which are hydrogen bonding through their side chains.** These asparagines are not conserved among AT sequences, which have a variety of hydrophilic side chains at equivalent positions.

have threonine at this position. Unlike the pectate lyases there are no regular "stacks" of identical residues packed in layers within the core, but a row of asparagine residues is found forming an extended hydrogen bonding chain on the surface of the β helix (Fig. 3). Overall the electron density for the model is very clear, but model building became increasingly difficult toward the C terminus. LAFIRE, a model fitting and rebuilding

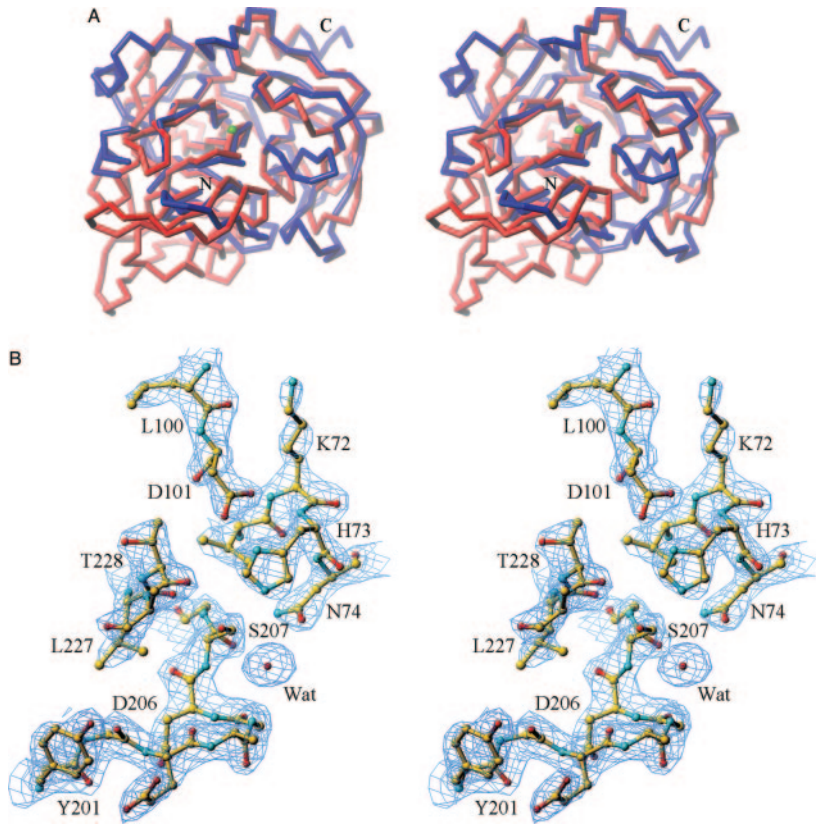

FIG. 4. *Top*, a stereo overlay of the Cα trace of bovine trypsin (*blue*) and the serine protease domain of Hbp (*red*). The Cα of the catalytic serine (Ser^{207}) is shown as a *green sphere*. Trypsin (PDB 5PTP) was matched to HBP by least-squares fitting residues 16–19 (1–4), 21–24 (15–18), 26–32 (49–55), 42–60 (58–76), 63–69 (78–84), 79–93 (85–99), 101–114 (100–113), 118–127 (114–123), 129–132 (128–131), 134–146 (136–148), 147–152 (156–161), 153–165 (163–175), 176–183 (177–184), 184–187 (186–189), 188–202 (200–214), 224–231 (236–243), and 234–245 (244–255). These 648 main-chain atoms selected by DALI have a root mean square deviation of 2.5 Å, and a maximum deviation of 8.2 Å. *Bottom*, the $2mF_o - DF_c$ electron density map contoured at 1.3σ over the catalytic site of the serine protease domain, showing the catalytic serine (Ser^{207}) and a water molecule hydrogen bonded to it. Tyr^{201}, the residue in the specificity pocket, is seen on the *lower left*.

program using local correlation of electron density,[2] was essential in assembling the last 100 residues of the model. Details of the program may be found at: altair.sci.hokudai.ac.jp/g6/Research/Lafire_English.html. There are several breaks in the final electron density maps in this region where short loops cannot be modeled. The C-terminal region from residue 950 to the end is a conserved domain found among AT proteins believed to assist folding after passage through the C-terminal pore domain (37).

Residues 1–256 of the model form a clearly separate domain (*domain 1* in Fig. 1), including six β strands rolled into a barrel-like structure, with several long β hairpins over its surface. The N-terminal glycine residue is deep within the structure, close to the bottom of the barrel. Folding of the domain can clearly only occur after the signal peptide has been removed. The glycine N atom hydrogen bonds to Asp^{206}, which also interacts with Arg^{140} and the main-chain of Gly^{144}. An α-helix (Leu^{245}-Asp^{255}) connects this domain to the β helical stem. Searching the PDB for structural homologues with DALI (38, 39) revealed that the N-terminal domain is highly similar to bovine trypsin (PDB 5PTP), which gave a Z-score of 17.5 (scores above 2.0 are considered significant). The 162 residues matched showed 20% sequence identity to Hbp and could be fitted with root mean square deviation of 2.5 Å over the main chain. The Cα traces are overlaid in Fig. 4, *top*. Other proteases gave rather poorer scores. Overlaying trypsin on the model of Hbp gave an excellent match of the catalytic triad (Ser^{195}, His^{57}, and Asp^{102}) of trypsin with Ser^{207}, His^{73}, and Asp^{101} of Hbp (Fig. 4, *bottom*). Comparing the active sites, Hbp misses a loop (Ser^{93} to Asp^{101} in trypsin) just prior to the essential aspartate. In trypsin this loop includes Leu^{99}, which contacts Trp^{215}, one of the residues that forms antiparallel-β sheet interactions with the substrate polypeptide. Trp^{215} in trypsin is changed to alanine (Ala^{229}) in Hbp, making the active site of Hbp much more open. At the bottom of the specificity pocket, which accommodates the residue preceding the scissile bond, trypsin has an aspartate (Asp^{189}), which leads to the preference for Lys and Arg residues at the S1 position. Chymotrypsin and elastase have serine in this position, but Hbp has tyrosine (Tyr^{201}). Among the trypsin/chymotrypsin family, Pro^{225} is very highly conserved from bacteria to man, but Hbp misses a loop just before this residue, and the nearest equivalent is Gly^{237}. Hbp is also unusual in having an asparagine residue at the position equivalent to glycine 226 in trypsin. Thus, although the fold of the N-terminal domain is close to trypsin, it also has unique features not seen in other serine proteases. The active site suggests broad substrate specificity with a preference for hydrophobic side chains in the specificity pocket. The very open active site presumably helps Hbp to attack folded globular proteins such as Hb, but more work is required to determine its functional properties as a protease. Low resolution electron density maps calculated using data collected with crystals soaked with two serine protease inhibitors, tosylphenylalanine chloromethyl ketone and tosyllysine chloromethyl ketone, show peaks of density close to the active site serine, suggesting the protease is still active at pH 4.6 (data not shown).

Sequence alignments of SPATE family members shows the sequence GDSGSPL to be highly conserved between Gly^{205} and Leu^{211} (Fig. 5). *vat*, an AT gene found in a pathogenicity island of avian pathogenic *E. coli* Ec222, has a slightly altered sequence, ATSGSPL, which may be related to its vacuolating activity (15), but it maintains both the active site serine (Ser^{207}) and Ser^{209}, which hydrogen bonds to the carbonyl oxygen of Asp^{206} to make the turn carrying Ser^{207}. The N-terminal Gly^{1} lies close to the active site serine and hydrogen bonds to the side chain of Asp^{206}. This strongly suggests the enzyme can only become active after the leader peptide is removed. All SPATE proteins tested show proteolytic activity, but with different substrate specificities, suggesting different

[2] M. Yao, unpublished program.

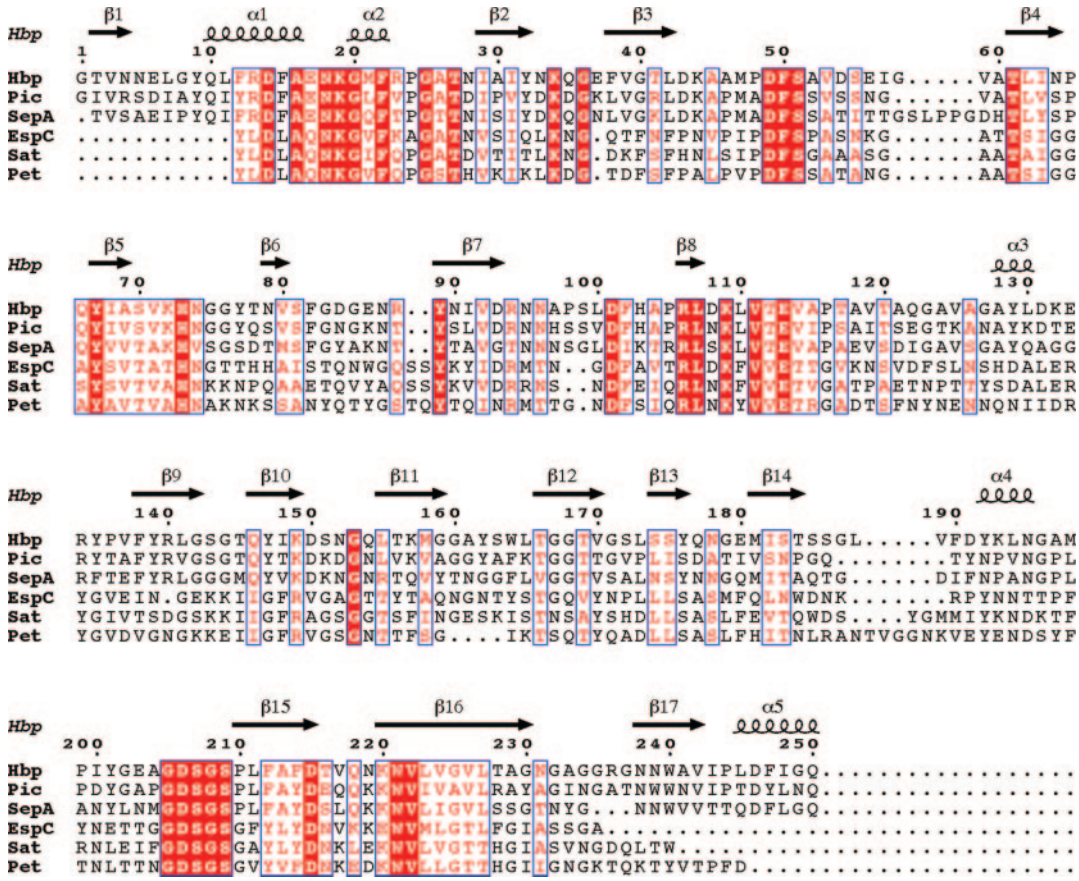

FIG. 5. **A sequence alignment of Hbp with other members of the SPATE family, covering the N-terminal protease domain.** Conservative mutations are shown in *red*, and residues common to all of the six compared sequences are shown in *white on red*. The UNIPROT reference numbers for the sequences are O88093 (*Hbp*), Q8CWC7 (*Pic*), Q8VSL2 (*SepA*), P77070 (*EspC*), Q6KD43 (*Sat*), and O68900 (*Pet*). Secondary structure elements of Hbp are shown over the alignment. The conserved regions show strong correlation to these structural elements, with the exception of the catalytic region GDSGS. The side-chain four residues upstream from this sequence sits in the specificity pocket (Tyr^{201} in the case of Hbp).

functions (9). Replacement of the active site serine of Pet (a SPATE from enteroaggregative *E. coli*) with isoleucine did not prevent release of the passenger domain from the C-terminal pore, but did eliminate all cytotoxic effects (40). Comparing the sequences of various ATs shows that Tyr^{201} in the specificity pocket of Hbp is preserved in Pic and SepA, whereas Pet and Sat have leucine (Fig. 5). The sequence alignment also suggests that EspC has glutamate at this position, which presumably would affect substrate preference strongly. Using the N domain structure of Hbp, a homology model of Pet protease domain was built using SwissModel (41). From this a second homology model of Sat protease domain was also constructed. The weaker sequence conservation between Hbp and Sat prevented a model of the latter being built more directly. These homology models show no unexpectedly loose packing or unfavorable geometry, indicating that each protein has identical connectivity, with minor differences in loop regions, in the N-terminal protease domain.

The N-terminal domain makes extensive contact with the β-helical stem of Hbp, but given its globular shape the contact region with the roughly cylindrical β helix is long and narrow. Many N domain contacts are found between Asp^{108} and Val^{114}. Arg^{106} makes buried hydrogen bonds to the carbonyl oxygen atoms of Asn^{301} and Gly^{303}. The main hydrophobic contact is provided by Leu^{110}, which is buried at the interface. Arg^{106} and Lys^{109} are very strongly conserved among ATs, the latter forming a salt bridge with Asp^{367}. Arg^{441} and Asp^{49}, both well conserved, make another bond between the N-terminal protease domain and the β-helix. These contacts, and the preservation of key glycines

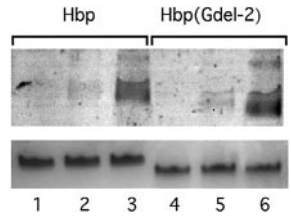

FIG. 6. **Heme binding by Hbp.** PCR was used to create an Hbp mutant (called "Gdel-2") missing domain 2 by linking Ala^{481} to Asn^{556} with a glycine residue. Purified native Hbp and Hbp(Gdel-2) (20 pmol) were incubated with different concentrations of heme and subsequently subjected to nondenaturing gel electrophoresis as described previously (22). Heme containing proteins were visualized by tetramethyl benzidine staining. The *upper panel* shows the tetramethyl benzidine-stained gel; *lanes 1–3*, Hbp incubated with 0, 50, or 100 pmol of heme, respectively. Heme binding, shown by the *dark band*, is unchanged in the mutant. *Lanes 4–6*, Hbp(Gdel-2) incubated with the different amounts of heme. The *lower panel* shows the Coomassie Blue-stained gel of Hbp and Hbp(Gdel-2). The lower molecular weight of the mutant is apparent.

in the sequences of ATs, suggest that these proteins share a very similar structure over the first half of the protein.

Residues 481–556 of Hbp make a small separate domain (*domain 2* in Fig. 1), the main chain returning to the β stem close to the point of departure. DALI found only one notable match, to the chitin binding domain of chitinase b (PDB 1E15). This domain consists of only 49 residues folded into three antiparallel β strands connected by long loops. The chitin binding domain fitted the 481–556 loop of Hbp with a Z-score of 3.6

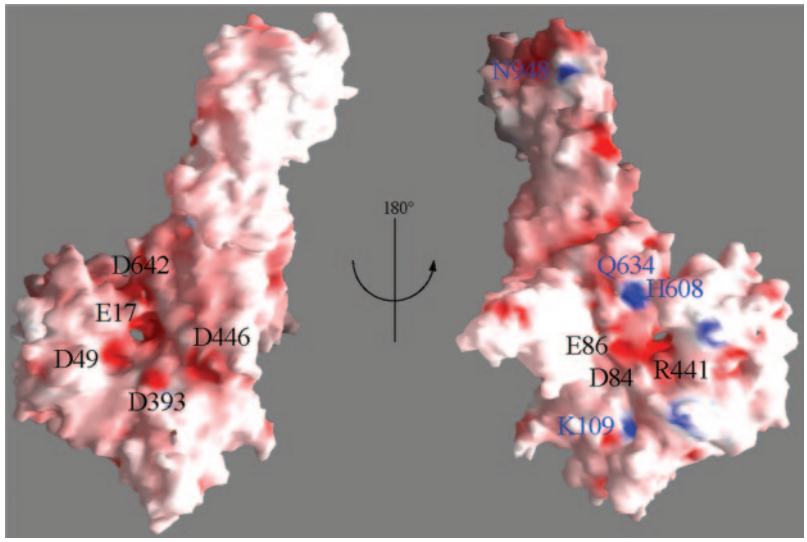

FIG. 7. **Surface charge of Hbp.** The molecular surface of the protein, colored by electrostatic charge (*red*, negative; *blue*, positive). *Color saturation* corresponds to an electron energy of $\pm 20\ k_BT$.

over 44 residues, with 16% sequence identity. Not surprisingly, the surface aromatic groups of chitin binding domain presumed to bind chitin are altered in Hbp, but the 481–556 loop has a different cluster of aromatic residues, which appear to form a binding pocket of some sort. Tyr^{510} and Tyr^{526} point into solvent and are highly exposed (and make crystal contacts with a neighboring molecule). The nearby Trp^{492}, Tyr^{506}, and Phe^{527} make a hydrophobic pocket open to solvent. The 481–556 domain is quite different to any known heme-binding site, and a mutant in which this domain is deleted shows identical heme binding to the wild-type protein (Fig. 6). Possibly its function is docking to extracellular matrix proteins or the bacterial receptor proteins, but these have yet to be identified. Using residues 250–615 from the Hbp model, a homology model of the same region could be built (again using SwissModel) for Pic. Key residues such as Glu^{538} and Arg^{554}, which hold domain 2 against the β-helix by a salt bridge, are preserved in Pic. Sequence alignment of ATs shows that Pet, Sat, and EspC among others do not have this domain.

Residues 608–644 of Hbp form a long discursive loop projecting from the β-helix, which also forms close contacts with the N domain. Thr^{616}-Ala^{626} forms an α helix, but this probably requires hydrophobic contact with the N domain residues Tyr^{9} and Met^{158} to form. The 608–644 loop has no hydrophobic core of its own to suggest it is a stable independent fold. Because other ATs show no homology to Hbp in this region, it may be the heme binding site. Residue 629 is histidine, which is the most common residue bonding to heme groups, but the structure does not readily suggest how heme might bind. Repeated attempts to co-crystallize Hbp with heme or soak the crystals were unsuccessful, and mutagenesis is now underway to test the function of this region. The charge distribution over the molecular surface shows a concentration of negative charge on one face of the molecule, centered around the interface between the β helix and domain 1 (Fig. 7). Further mutagenesis is required to determine whether this negatively charged patch is involved in heme transport.

Sequence similarity among ATs shows an increase again toward the C-terminal region of the passenger proteins. Glycines 868 and 875 in turns of Hbp are completely preserved, as are Met^{886} and Trp^{891}. The β-helix structure is broken at the conserved Ala^{979}/Pro^{980}, which mark a change in the direction of the polypeptide chain. Breaks in the electron density prevented modeling residues 913–918, 937–941, and 1024–1031. Despite the sequence conservation over the last 140 residues, these gaps in the Hbp model prevented making homology models of this region. The pattern of preserved residues matches very well interactions seen in the Hbp structure, however, and the C-terminal region of other ATs presumably shows a very similar fold. Key residues include Trp^{1015}, which is buried and forms a hydrogen bond through its indole side chain to make a turn in the protein backbone.

Hbp has been shown to cross the inner membrane by the signal recognition particle (SRP)-dependent pathway (3), but the precise mechanism of transport across the outer membrane of the cell remains unclear. Several different proposals have been made, including the suggestion that autotransporters form a multimeric pore in the outer membrane (42). Recently the structure of the pore-forming domain of NalP has been solved, showing that it is in fact a monomer, a 12-stranded β-barrel blocked by an N-terminal α helix (7). The pore diameter, roughly 10–12 Å, is insufficient to allow passage of folded proteins and suggests that only extended polypeptides can pass through. This appears incompatible with the idea that the passenger domain is threaded through from its C-terminal end (the hairpin model), because the pore would then have to accommodate two stretches of polypeptide simultaneously. An alternative model is that the N terminus of the passenger domain is threaded through the pore, but, because other protein domains fused to the pore domain are also transported, there is no evidence of sequence-specific interactions that target the N terminus to the pore. Oomen and colleagues (7) have suggested that the passenger domain may instead pass though the pore of the Omp85 complex, and that the name "autotransporter" is in fact a misnomer. From the crystal structure of Hbp described here, it is clear that the C-terminal domain (residues 950–1048) of the mature protein is more flexible and mobile than the N-terminal protease domain. This is consistent with the prediction that the C-terminal domain plays a key role in passage across the outer membrane. On the other hand, because the N-terminal glycine of the mature protein is buried within the structure, the protease domain cannot fold properly until the signal peptide is removed, and this may provide a mechanism for keeping the protein in a readily transported form.

Tsh (temperature-sensitive hemagglutination), a SPATE found in avian pathogenic *E. coli*, has only two amino acid changes compared with Hbp (14, 43, 44), but Hbp showed no hemagglutination activity in tests (22). Recently this result has been confirmed for Tsh (45). The two mutations, $Lys^{157} \rightarrow$ Gln and $Thr^{790} \rightarrow$ Ala, are both on the protein surface. Lys^{157} is roughly 18 Å from the active site, but replacement with gluta-

mine seems unlikely to alter the functional behavior of the protein greatly. Hbp and Tsh are therefore essentially identical, suggesting that Tsh too is a hemoglobin protease. If so, then the name "Tsh" seems redundant, and we suggest the name "Hbp" be used exclusively in the future. Tsh has been shown to bind red blood cells, hemoglobin, fibronectin, and collagen IV (45). It remains to be tested which of these are bound by domain 2.

DISCUSSION

Hbp was first purified from *E. coli* EB1, a strain isolated from an intra-abdominal wound infection (22). Without Hbp, *E. coli* and *B. fragilis* are unable to form abscesses and are more readily cleared from the site of infection by the immune system, presumably due to iron limitation. Although a great deal is known about iron-uptake by pathogens through siderophores, very few bacterial heme uptake systems have been characterized in molecular detail. The best understood is the HasA protein of *Serratia marcescens* for which a crystal structure has been refined to 1.9 Å (46). Hbp has a quite different fold, and is over five times larger. Much of the biology involving Hbp remains to be elucidated. No surface protein from either *E. coli* or *B. fragilis* has yet been identified that can interact with Hbp and receive heme from it, yet these must presumably exist given the known effects of the protein on these bacteria under conditions of iron limitation. *B. fragilis* up-regulates expression of a number of proteins under iron-limiting conditions (47), and these are currently being examined as possible receptors for Hbp.

One functional role of Hbp is to remove heme from Hb and supply it to bacteria growing within the body. The crystal structure shows a novel trypsin-like serine protease domain that is well conserved among the SPATE family. The overall β-helix architecture is also common to the ATs, but different loops and domains carried by this stable fold appear to give the proteins distinctive functional properties. Hbp does not have a sequence motif or structural domain, which resembles a known heme binding site. Further heme binding studies with mutants of Hbp and other ATs are currently underway to determine how heme interacts with Hbp.

Homology models of other ATs based on the Hbp structure described here will hopefully be useful in designing new experiments to test their functional roles and find ways to block their activity. Hbp may be a convenient vehicle for presenting peptides for antibody and vaccine production, by inserting them in place of domain 2. Domain 2 itself may well allow us to produce a clinically useful vaccine to reduce substantially the occurrence of severe peritonitis.

Acknowledgments—We Drs. N. Kamiya, Y. Kawano, and H. Nakajima of SPring8 and Prof. S. Wakatsuki and Drs. M. Suzuki, N. Matsunaga, and N. Igarashi of the Photon Factory for help with data collection.

REFERENCES

1. Henderson, I. R., Cappello, R., and Nataro, J. P. (2000) *Trends Microbiol.* **8,** 529–532
2. Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'Aldeen, D. (2004) *Microbiol. Mol. Biol. Rev.* **68,** 692–744
3. Sijbrandi, R., Urbanus, M. L., ten Hagen-Jongman, C. M., Bernstein, H. D., Oudega, B., Otto, B. R., and Luirink, J. (2003) *J. Biol. Chem.* **278,** 4654–4659
4. Klauser, T., Pohlner, J., and Meyer, T. F. (1993) *Bioessays* **15,** 799–805
5. Pohlner, J., Halter, R., Beyreuther, K., and Meyer, T. F. (1987) *Nature* **325,** 458–462
6. Jose, J., Jahnig, F., and Meyer, T. F. (1995) *Mol. Microbiol.* **18,** 378–380
7. Oomen, C. J., Van Ulsen, P., Van Gelder, P., Feijen, M., Tommassen, J., and Gros, P. (2004) *EMBO J.* **23,** 1257–1266
8. Henderson, I. R., and Nataro, J. P. (2001) *Infect. Immun.* **69,** 1231–1243
9. Dutta, P. R., Cappello, R., Navarro-Garcia, F., and Nataro, J. P. (2002) *Infect. Immun.* **70,** 7105–7113
10. Stein, M., Kenny, B., Stein, M. A., and Finlay, B. B. (1996) *J. Bacteriol.* **178,** 6546–6554
11. Brunder, W., Schmidt, H., and Karch, H. (1997) *Mol. Microbiol.* **24,** 767–778
12. Eslava, C., Navarro-Garcia, F., Czeczulin, J. R., Henderson, I. R., Cravioto, A., and Nataro, J. P. (1998) *Infect. Immun.* **66,** 3155–3163
13. Guyer, D. M., Henderson, I. R., Nataro, J. P., and Mobley, H. L. (2000) *Mol. Microbiol.* **38,** 53–66
14. Provence, D. L., and Curtiss, R., 3rd. (1994) *Infect. Immun.* **62,** 1369–1380
15. Parreira, V. R., and Gyles, C. L. (2003) *Infect. Immun.* **71,** 5087–5096
16. Henderson, I. R., Czeczulin, J., Eslava, C., Noriega, F., and Nataro, J. P. (1999) *Infect. Immun.* **67,** 5587–5596
17. Benjelloun-Touimi, Z., Si Tahar, M., Montecucco, C., Sansonetti, P. J., and Parsot, C. (1998) *Microbiology* **144,** 1815–1822
18. Farthmann, E. H., and Schoffel, U. (1998) *Infection* **26,** 329–334
19. Aldridge, K. E. (1995) *Am. J. Surg.* **169,** 2S–7S
20. Rotstein, O. D., Kao, J., and Houston, K. (1989) *J. Med. Microbiol.* **29,** 269–276
21. Otto, B. R., van Dooren, S. J., Dozois, C. M., Luirink, J., and Oudega, B. (2002) *Infect. Immun.* **70,** 5–10
22. Otto, B. R., van Dooren, S. J., Nuijens, J. H., Luirink, J., and Oudega, B. (1998) *J. Exp. Med.* **188,** 1091–1103
23. Williams, P. H. (1979) *Infect. Immun.* **26,** 925–932
24. Tame, J. R. H., van Dooren, S. J., Oudega, B., and Otto, B. R. (2002) *Acta Crystallogr. D. Biol. Crystallogr.* **58,** 843–845
25. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
26. Terwilliger, T. C. (2003) *Methods Enzymol.* **374,** 22–37
27. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D. Biol. Crystallogr.* **55,** 849–861
28. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119
29. Roussel, A., and Cambillau, C. (1989) *TURBO-FRODO*, Silicon Graphics, Mountain View, CA
30. Brunger, A. T. (1996) *X-PLOR version 3.85*, Yale University Press, New Haven, CT
31. CCP4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
32. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55,** 247–255
33. Emsley, P., Charles, I. G., Fairweather, N. F., and Isaacs, N. W. (1996) *Nature* **381,** 90–92
34. Clantin, B., Hodak, H., Willery, E., Locht, C., Jacob-Dubuisson, F., and Villeret, V. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6194–6199
35. Jenkins, J., and Pickersgill, R. (2001) *Prog. Biophys. Mol. Biol.* **77,** 111–175
36. Scott, D. J., Grossmann, J. G., Tame, J. R., Byron, O., Wilson, K. S., and Otto, B. R. (2002) *J. Mol. Biol.* **315,** 1179–1187
37. Oliver, D. C., Huang, G., Nodel, E., Pleasance, S., and Fernandez, R. C. (2003) *Mol. Microbiol.* **47,** 1367–1383
38. Holm, L., and Sander, C. (1995) *Trends Biochem. Sci.* **20,** 478–480
39. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
40. Navarro-Garcia, F., Canizalez-Roman, A., Luna, J., Sears, C., and Nataro, J. P. (2001) *Infect. Immun.* **69,** 1053–1060
41. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) *Nucleic Acids Research* **31,** 3381–3385
42. Veiga, E., Sugawara, E., Nikaido, H., de Lorenzo, V., and Fernandez, L. A. (2002) *EMBO J.* **21,** 2122–2131
43. Stathopoulos, C., Provence, D. L., and Curtiss, R., 3rd. (1999) *Infect. Immun.* **67,** 772–781
44. Dozois, C. M., Dho-Moulin, M., Bree, A., Fairbrother, J. M., Desautels, C., and Curtiss, R., 3rd. (2000) *Infect. Immun.* **68,** 4145–4154
45. Kostakioti, M., and Stathopoulos, C. (2004) *Infect. Immun.* **72,** 5548–5554
46. Arnoux, P., Haser, R., Izadi, N., Lecroisey, A., Delepierre, M., Wandersman, C., and Czjzek, M. (1999) *Nat. Struct. Biol.* **6,** 516–520
47. Otto, B. R., Verweij-van Vught, A. M., van Doorn, J., and Maclaren, D. M. (1988) *Microb. Pathog.* **4,** 279–287